Plant Cell Rep (2008) 27:1645–1653
DOI 10.1007/s00299-008-0568-z

GENETICS AND GENOMICS

# ATG-anchored AFLP (ATG-AFLP) analysis in cotton

**Yingzhi Lu · Jessica Curtiss · Danielle Miranda · Ed Hughs · Jinfa Zhang**

Received: 8 February 2008 / Revised: 20 May 2008 / Accepted: 26 May 2008 / Published online: 27 June 2008


**Abstract** Amplified fragment length polymorphism (AFLP) marker system has had broad applications in biology. However, the anonymous AFLP markers are mainly amplified from non-coding regions, limiting their usefulness as a functional marker system. To take advantages of the traditional AFLP techniques, we propose substitution of a restriction enzyme that recognizes a restriction site containing ATG, called ATG-anchored AFLP (ATG-AFLP) analysis. In this study, we chose *Nsi*I (recognizing ATG-CAT) to replace *Eco*RI in combination with *Mse*I to completely digest genomic DNA. One specific adaptor, one pre-selective primer and six selective amplification primers for the *Nsi*I site were designed for ligation and PCR. Six *Nsi*I and eight *Mse*I primers generated a total of 1,780 ATG-AFLP fragments, of which 750 (42%) were polymorphic among four genotypes from two cultivated cotton species (Upland cotton, *Gossypium hirsutum* and Pima cotton, *G. barbadense*). The number of ATG-AFLP markers was sufficient to separate the four genotypes into two groups, consistent with their evolutionary and breeding history. Our results also showed that ATG-AFLP generated less number of total and polymorphic fragments per primer combination (2–3 vs. 4–5) than conventional AFLP within Upland cotton. Using a recombination inbred line (RIL) population, 62 polymorphic ATG-AFLP markers were mapped to 19 linkage groups with known chromosome anchored simple sequence repeat (SSR) markers. Of the nine ATG-AFLP fragments randomly chosen, three were found to be highly homologous to cotton cDNA sequences. An in-silico analysis of cotton and *Arabidopsis* cDNA confirmed that the ATG-anchored enzyme combination *Nsi*I/*Mse*I did generate more fragments than the *Eco*RI/*Mse*I combination.

**Keywords** *Gossypium hirsutum* · *G. barbadense* · AFLP · ATG-AFLP

Communicated by C. Quiros.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00299-008-0568-z) contains supplementary material, which is available to authorized users.

Y. Lu · J. Curtiss · D. Miranda · J. Zhang (✉)
Department of Plant and Environmental Sciences,
New Mexico State University, Las Cruces, NM 88003, USA
e-mail: jinzhang@nmsu.edu

E. Hughs
USDA-ARS, Southwestern Cotton Ginning Research Lab,
Mesilla Park, NM 88003, USA

## Introduction

Since its invention (Vos et al. 1995), the amplified fragment length polymorphism (AFLP) marker system has found its applications in a wide range of research areas including biology, genetics, breeding, evolution, and ecology (Bensch and Akesson 2005; Meudt and Clarke 2007; Vuylsteke et al. 2007a). For example, AFLP has been used in cotton for linkage map construction (Altaf et al. 1998; Brubaker and Brown 2003; Lacape et al. 2003; Mei et al., 2004), gene mapping (Lacape et al. 2005; Niu et al. 2007; Wang et al. 2007; Zhang et al. 2005b), germplasm diversity assessment (Abdalla et al. 2001; Iqbal et al. 2001; Pillay and Myers 1999; Westengen et al. 2005; Zhang et al. 2007), and evolutionary study (Liu et al. 2001).

The AFLP analysis combines the reliability of restriction enzyme digestion with the utility of the polymerase chain reaction. AFLP detects the presence of point mutations, insertions, deletions and other genetic rearrangements, and is very reproducible and reliable. Because complete genome

sequence information is still not available for most organisms of agronomic or biological importance, AFLP has become the marker system of choice due to the no need for prior sequence information and its high multiplexity and hence high throughput nature in marker data production per primer combination. To improve its discriminatory power, reproducibility, and ease of interpretation and standardization, various modifications to the original AFLP techniques have been proposed that have involved restriction enzyme combinations including using one enzyme, primer extensions (1, 2, or 3 bp) and combinations (with other primers based on SSR, retroposons, and disease resistance gene analogues), and fragment separation systems (e.g. Park et al. 2003; Roy et al. 2002; Zhang et al. 2007). Van der Wurff et al. (2000) proposed TE-AFLP using three enzymes before ligation, which was shown to reduce the number of AFLP bands and increase discriminatory power. Recently, Zhang et al. (2005a) showed that further restriction of AFLP products by a third restriction enzyme can release many more polymorphic fragments, called cleaved AFLP (cAFLP) analysis.

However, as with any other marker systems, AFLP is not without disadvantages. Even though AFLP markers are widely distributed throughout a genome, they are often found to be clustered or concentrated in centromeric regions (Meudt and Clarke 2007). The anonymous AFLP markers are a sampling of most non-coding sequences in the genome (Meudt and Clarke 2007). To genome-widely target coding sequences, herein we report a modified AFLP analysis, called ATG-anchored AFLP (ATG-AFLP) analysis using cotton as an example.

## Material and methods

### Plant materials

Four tetraploid cotton genotypes, i.e. TM-1, Acala 1517-99, NM 24016, and Pima 3-79, were used in this study. TM-1 is the genetic standard for Upland cotton (Kohel et al. 1970), while Pima 3-79 is the genetic standard for *G. barbadense* (Endrizzi et al. 1984). Acala 1517-99 was an Acala cotton cultivar released from New Mexico State University (Cantrell et al. 2000) and NM 24016 was an Acala breeding line with substantial germplasm introgression from *G. barbadense* (Cantrell and Davis 2000). DNA was extracted from cotton young leaves using the method of Zhang and Stewart (2000).

A recombinant inbred line (RIL) mapping population was made from the cross of Pima 3-79 as the female and NM 24016 as the male. The resulted $F_1$ was grown in the greenhouse for selfing to generated $F_2$ seed. The $F_2$ and the following generations were grown in the same greenhouse for generation advancement using single seed descent (SSD). A total of 60 individual $F_7$ plants were sampled for DNA extraction using the above-mentioned method.

### Design of *Nsi*I-adaptor and *Nsi*I primers

*Nsi*I, which recognizes 5′ATGCAT3′, was used in combination with *Mse*I (recognizing 5′TTAA3′) in this study. The *Nsi*I-adaptor was modified from *Eco*RI adaptor (Applied Biosystem Inc., Foster City, CA) with changes on several nucleotide positions in that, (1) the cohesive end was compatible with *Nsi*I; (2) ligation of the adaptor to the restricted DNA would not regenerate the recognition site for the same enzyme; and (3) the GC content of the adaptor is between 40 and 50%. The sequence of the *Nsi*I-adaptor is,

5′-GACTGCGTACTTGCA-3′
3′-CTGACGCATGA-5′.

For pre-selective PCR amplifications (PSA), the PSA primer for the *Nsi*I adaptor is, forward *Nsi*I-PSA: 5′-GACTGCGTACTTGCA-3′, while the PSA primer for *Mse*I is, reverse *Mse*I-PSA: 5′-GATGAGTCCTGAGTAAC-3′, the same as previously reported (Applied Biosystem Inc., Foster City, CA, USA). AFLP primers for selective amplification (SA) consist of the adaptor sequence and three selective nucleotides. Pre-selective amplification using the AFLP primers based on the adaptor sequence with one selective nucleotide is strongly recommended for plant species with a large genome size such as cotton. In this study we selected "A" as the first selective nucleotide for PSA *Nsi*I primer (forward). The second and third selective nucleotides were selected with consideration of the genetic coding characteristics of three nucleotides. For example, both "AAT" and "AAC" code for Thr. Only "C" was chosen since it gives the primer a higher GC content with similar annealing temperature to that of SA primer for *Mse*I sites. "ATG" was excluded because it was contained in the *Nsi*I recognition site. Six *Nsi*I selective primers were designed (Table 1) and used as forward primers. The sequences of *Mse*I-adaptor, *Mse*I pre-selective primer, and eight selective primers were the same as these based on the protocol of Applied Biosystem Inc.

### AFLP protocol

The AFLP procedure was performed following the protocol of Vos et al. (1995) with minor modifications. One mg of genomic DNA was restricted using 10 U *Nsi*I and *Mse*I. The pre-selective PCR amplifications were performed using pre-selective primers (*Nsi*I + A, *Mse*I + C), followed by selective PCR using six *Nsi*I primers and eight *Mse*I primers. The protocol was the same as that detailed in Zhang et al. (2005a) except that the primer concentrations were 1 μM for *Nsi*I primers and 1.25 μM for *Mse*I primers. Six

**Table 1** Sequences of *Nsi*I selective primers used in ATG-AFLP

| Primer | Sequence |
|---|---|
| N1 | 5′-CTGCGTACTTGCATAAC-3′ |
| N2 | 5′-CTGCGTACTTGCATAAG-3′ |
| N3 | 5′-CTGCGTACTTGCATACT-3′ |
| N4 | 5′-CTGCGTACTTGCATATC-3′ |
| N5 | 5′-CTGCGTACTTGCATAGA-3′ |
| N6 | 5′-CTGCGTACTTGCATAGT-3′ |

*Nsi*I and eight *Mse*I primers were used to make a total set of 48 *Nsi*I-*Mse*I combinations (Table 2). An aliquot of 5 µl AFLP PCR reaction was loaded on a 5% polyacrylamide gel for fragment separation. After 1.5 h of electrophoresis at a constant 50 Watts, the gel was visualized by silver staining modified based on Promega's protocol.

In order to compare the polymorphic rate detected by ATG-AFLP, three *Eco*RI-*Mse*I primer combinations were used to screen TM-1 and NM 24016 using the same the PCR protocols and gel electrophoresis techniques as described for ATG-AFLP.

## Cloning and sequencing of ATG-AFLP fragments

Polymorphic ATG-AFLP bands were identified, cut and introduced into a clean tube containing 100 µl of $ddH_2O$. The gel slice was stored at 4°C overnight before use. To obtain enough DNA for cloning, a re-amplification step was employed by using 2–5 µl elute as the DNA template in a PCR reaction with the same primers and the PCR conditions. The re-amplified DNA fragment was recovered from the agarose gel using Qiagen Gel Extraction Kit (Valencia, CA, USA). The purified DNA was cloned into the pGEM-T Easy Vector system following Promega's protocol (Madison, WI, USA). The sequencing was performed by a Li-Cor DNA Sequencer at the Department of Chemistry and Biochemistry, New Mexico State University (Las Cruces, NM, USA).

## ATG-AFLP marker scoring and mapping

Forty-eight *Nsi*I-*Mse*I primer pairs were used to screen the above-mentioned four cotton genotypes. Presence or absence of each ATG-AFLP fragment was scored as a binary unit character (i.e., 1 = present, 0 = absent). Genetic similarities based on Jaccard's coefficient were calculated using the Numerical Taxonomy Multivariate Analysis System (NTSYS-pc) version 2.1 software. The resulting genetic similarity matrices were used to generate an unweighted pair group of arithmetic means (UPGMA) dendrogram. The robustness of branching points was validated using bootstrapping with 500 permutations using an open source software, FreeTree (http://www.natur.cuni.cz/~flegr/programs/freetree.htm; Hampl et al. 2001). The tree was then drawn using Tree View (http://taxonomy.oology.gla.zac.uk/rod/treeview.html).

Eight primer pairs for ATG- AFLP were used to perform PCR reactions using the RIL population mentioned earlier. The two parents were included to evaluate the origin of the polymorphic bands. Each polymorphic band was scored as a single dominant marker. Data were analyzed with Mapmaker version 3.0 for PC (Lander et al. 1987). A minimum LOD score of 3.0 and a recombination fraction of 0.40 were used for map construction.

## In-silico analysis

To compare the modified ATG-anchored AFLP with the traditional AFLP, the software AFLPinSilico (http://bioinformatics.psb.ugent.be/webtools/aflpinsilico; Rombauts

**Table 2** The number of total (T) and polymorphic ATG-AFLP (PM) fragments amplified by 48 *Nsi*I (N1–N6)-*Mse*I (M1–M8) primer combinations

| | N1 | | N2 | | N3 | | N4 | | N5 | | N6 | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PM[a] | T[b] | PM | T | PM | T | PM | T | PM | T | PM | T | PM | T |
| M1 | 12 | 32 | 18 | 46 | 18 | 41 | 18 | 44 | 17 | 39 | 18 | 31 | 101 | 233 |
| M2 | 21 | 33 | 17 | 33 | 20 | 43 | 16 | 31 | 21 | 33 | 14 | 29 | 109 | 202 |
| M3 | 12 | 51 | 11 | 40 | 5 | 27 | 8 | 38 | 7 | 39 | 19 | 50 | 62 | 245 |
| M4 | 31 | 51 | 20 | 33 | 31 | 50 | 12 | 36 | 4 | 45 | 25 | 50 | 123 | 265 |
| M5 | 28 | 58 | 17 | 49 | 16 | 48 | 8 | 50 | 23 | 57 | 21 | 51 | 113 | 313 |
| M6 | 7 | 39 | 7 | 36 | 8 | 23 | 8 | 28 | 7 | 17 | 6 | 24 | 43 | 167 |
| M7 | 8 | 23 | 15 | 32 | 23 | 34 | 20 | 34 | 21 | 32 | 13 | 28 | 100 | 183 |
| M8 | 16 | 25 | 19 | 33 | 24 | 35 | 9 | 18 | 13 | 26 | 18 | 35 | 99 | 172 |
| Total | 135 | 312 | 124 | 302 | 145 | 301 | 99 | 279 | 113 | 288 | 134 | 298 | 750 | 1,780 |

[a] Polymorphic

[b] Total

et al. 2003) was used to generate cDNA-AFLP fragments for cotton unigenes constructed from Upland cotton EST sequences (http://plantta.tigr.org/cgi-bin/plantta_release.pl; Childs et al. 2007). Since there is no complete genome sequence available for cotton, *Arabidopsis thaliana* genome and cDNA sequences were downloaded from TAIR datasets (http://www.arabidopsis.org) for comparing the numbers of AFLP fragments from genes and intergenic regions generated by *Eco*RI/*Mse*I and *Nsi*I/*Mse*I using the software mentioned above. Genome sequences from yeast chromosome I, III and VI were also used to compare the numbers of AFLP fragments generated by *Eco*RI/*Mse*I and *Nsi*I/*Mse*I using the online software (http://insilico.ehu.es/AFLP; Bikandi et al. 2004).

## Results

### ATG-AFLP among Upland cotton, Pima cotton and Acala cotton

The 48 combinations of six *Nsi*I selective primers and eight *Mse*I selective primers generated a total of 1,780 fragments from the four cotton genotypes. The majority of PCR bands were between 100 and 600 bp. Of the 1,780 PCR fragments, 750 (42%) were polymorphic with an average of 15–16 per primer pair (Table 2). Different primer combinations produced varied numbers of total amplified fragments and polymorphic fragments. Primer pair N3M4 detected the highest level of polymorphism (31 polymorphic ATG-AFLP bands out of a total of 50), while N5M4 the lowest level of polymorphism (only four polymorphic bands out of a total of 45).

The polymorphism between Upland cotton and Pima cotton was assayed by a comparison between the genetic standards TM-1 and Pima 3-79. Of a total of 688 ATG-AFLP fragments, 351 of TM-1 origin and 337 of 3-79 origin were polymorphic (Table 3). Acala 1517-99 has 25 ATG-AFLP bands of the Pima 3-79 type, while NM 24016 has 88 Pima type bands (Table 3). Interestingly, both Acala 1517-99 and NM 24016 showed new polymorphic bands, i.e. 24 unique to Acala 1517-99, 16 unique to NM 24016, and 22 Acala cotton type bands shared by both (perhaps of the Pima cotton origin).

### Polymorphism within Upland cotton based on ATG-AFLP

The numbers of polymorphic bands among Upland cotton genotypes were 69 between TM-1 and Acala 1517-99, 126 between TM-1 and NM 24016, and 119 between Acala 1517-99 and NM 24016. The highest level of polymorphism (7%) within Upland cotton was found by a comparison between NM 24016 and TM-1. On average each primer pair amplified 2–3 polymorphic bands between TM-1 and NM 24016.

**Table 3** Distribution of ATG-AFLP fragments categorized by genotypes

| | TM1 type | Pima 3-79 type | Acala type | Acala 1517-99 type | NM 24016 type |
|---|---|---|---|---|---|
| TM1 | 351 | – | – | – | – |
| Pima 3-79 | – | 337 | – | – | – |
| Acala 1517-99 | 665 | 23 | 22 | 24 | – |
| NM 24016 | 599 | 89 | 22 | – | 16 |

As a comparison, a high level of polymorphism between TM-1 and NM 24016 was also detected by the commonly used AFLP system using *Eco*RI and *Mse*I restriction enzymes. Randomly using three primer pairs E1/M1, E2/M2, and E3/M3, 13 (8.8%) polymorphic bands (4–5 bands per primer pair) were identified from a total of 148 AFLP fragments. Using a higher resolution electrophoresis system, CEQ 8000 capillary sequencer, 15 polymorphic fragments were identified with the same three primer pairs. The results showed that the higher resolution of capillary DNA sequencer over the manual sequencer and silver staining techniques did not significantly benefit the detection of polymorphism. The lower number of total fragments and polymorphic fragments amplified by ATG-AFLP primers indicates that the *Nsi*I recognition sequence of ATGCAT occurred less frequently than the *Eco*RI recognition site (GAATTC) in the cotton genome.

### Cluster analysis using ATG-anchored AFLP markers

The genetic similarity matrix of ATG-AFLP data for the four cotton genotypes was constructed based on Jaccard's similarity coefficients (Table 4). As expected, the genetic similarities between Pima 3-79 and the three Upland cotton genotypes were low (ranging from 59.35 to 62.99%), while the three Upland cotton genotypes had higher similarity (>90%). TM-1 and Acala 1517-99 had the highest similarity (95.19%). The genetic similarities between NM 24016 and TM-1, and Acala 1517-99 were similar, 91.46 and 91.95%, respectively.

**Table 4** Similarity coefficients between four tetraploid cotton genotypes based on ATG-anchored AFLP markers

| | TM-1 | Acala 1517-99 | NM 24016 | Pima 3-79 |
|---|---|---|---|---|
| TM-1 | 1 | | | |
| Acala 1517-99 | 0.9519 | 1 | | |
| NM 24016 | 0.9146 | 0.9195 | 1 | |
| Pima 3-79 | 0.6025 | 0.5935 | 0.6299 | 1 |

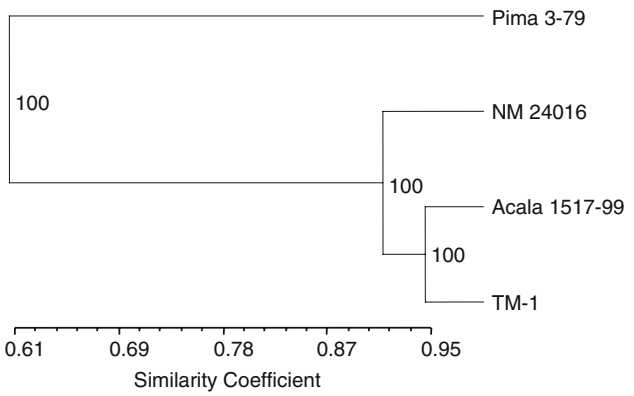


**Fig. 1** Cluster analysis of four tetraploid cotton genotypes based on ATG-AFLP marker data

The genetic similarity coefficients for the four cotton genotypes were then used to generate a phylogenetic dendrogram using the UPGMA method (Fig. 1). TM-1 and Acala 1517-99 were grouped first because of the highest genetic similarity. NM 24016 was not grouped with TM-1 and Acala 1517-99, reflecting a relatively higher genetic distance from these two Upland cotton genotypes. NM 24016 was then joined with TM-1 and Acala 1517-99 to form the Upland cotton group. As expected, Pima cotton 3-79 formed a separate group.

## Segregation of ATG-AFLP markers in RIL population

A total of eight combinations of *Nsi*I and *Mse*I selective primers were selected to screen the RIL population made from a cross between 3-79 and NM 24016 to examine the specificity and efficiency of ATG-AFLP marker system in cotton plants. After six generations of selfing from the $F_1$ of 3-79 x NM 24016, the progeny plants displayed the segregation of the polymorphic AFLP bands, as expected (Fig. 2). The majority of ATG-AFLP markers obtained from the parental screening was also detected in the tested RIL population samples, while polymorphism for some ATG-AFLP fragments disappeared in the segregating population.

From the eight primer pairs, a total of 62 ATG-AFLP polymorphic markers were scored from the RIL population of 60 lines (data not shown). Thirty-six ATG-AFLP markers segregated normally (1:1 ratio) as expected, while 26 displayed significant segregation distortions (SD) ($P < 0.05$). Of the 26 SD markers, 14 were biased toward to the female parent, Pima 3-79, while 12 were preferentially biased toward to the male parent NM 24016.

## Linkage maps of ATG-anchored AFLP markers

A majority of the ATG-AFLP markers (52 out of 62), were linked with 40 SSR markers in 22 groups (Fig. 3).

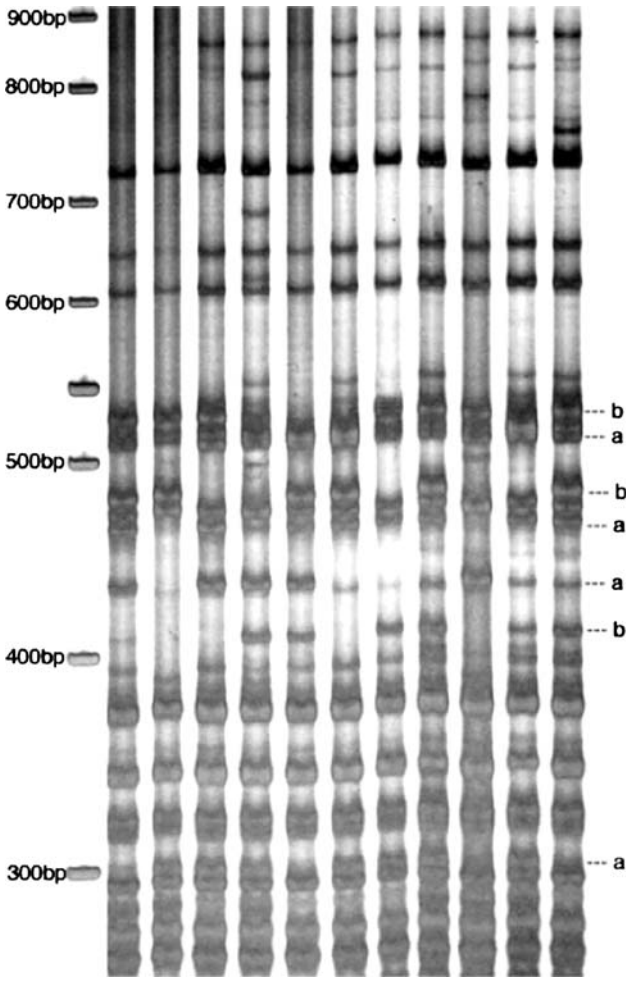


**Fig. 2** ATG-AFLP fragments separated on a 5% polyacrylamide gel from a RIL population of Pima 3-79 x NM 24016. **a** Pima 3-79 type; **b** NM 24016 type

Seventeen linkage groups containing 39 ATG-AFLP markers were assigned to 14 chromosomes using the 40 chromosome-anchored SSR markers. Thirteen unassigned ATG-AFLP markers were left to form 5 linkage groups. Interestingly, 13 SD ATG-AFLP markers favoring Pima 3-79 were clustered on chromosome c15, implying the existence of a reproductive barrier region on this chromosome. There was only one SD marker favoring Pima 3-79 located on chromosome c5. In contrast, 11 SD markers favoring Upland cotton NM 24016 were spread onto chromosome c2, c6 (or c17), c10, c26 and other unassigned groups.

## Cloning and sequencing of ATG-AFLP fragments

Nine AFLP fragments were randomly chosen, cloned and sequenced to confirm the specificity of ATG-AFLP amplifications. All the 9 cloned fragments contained the exact sequences of forward and reverse primers (data not shown). Four ATG-AFLP marker sequences showed high

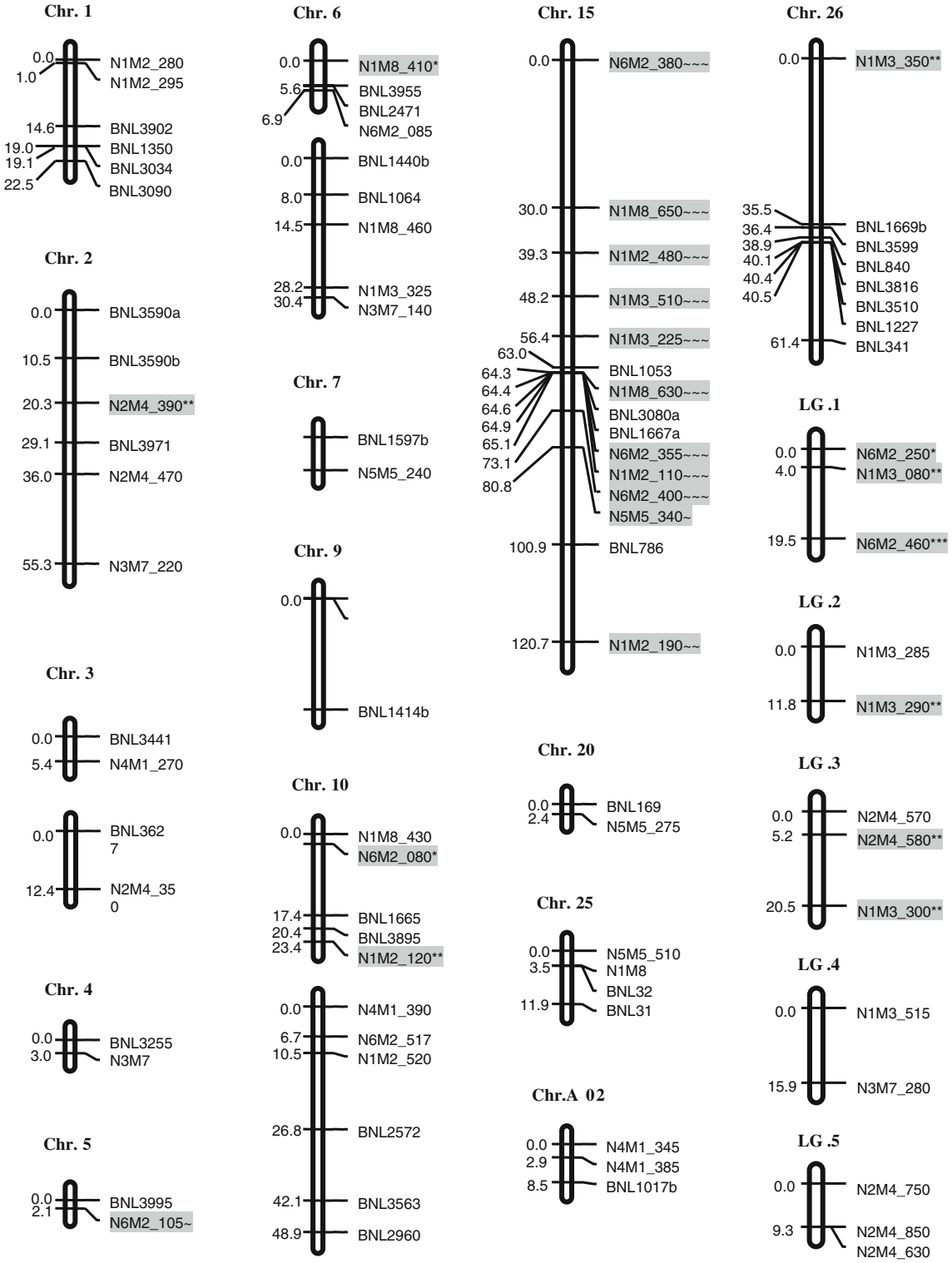


**Fig. 3** Genetic linkage map of cotton constructed from 60 recombinant inbred lines (RILs) derived from the interspecific cross *Gh cv*. NM 24016 x *Gb cv*. Pima 3-79 using 62 ATG-AFLP and framework SSR markers to anchor the map. The markers that show significant segregation distortion are indicated by *gray blocks* with symbols showing an excess of either *Gh* (*) or *Gb* (∼) genotypes (* and ∼, $P < 0.05$; ** and ∼∼, $P < 0.01$; *** and ∼∼∼, $P < 0.001$). Map distances are given in cM using the Kosambi mapping function

**Table 5** Homology of ATG-AFLP marker sequences

| Marker name | Origin | Sequence length (bp) | Homologous sequence in GenBank | | | |
|---|---|---|---|---|---|---|
| | | | Accession | Gene name | Identity | *E* value |
| N1M2_295 | Pima 3-79 | 274 | No homology | | | |
| N1M2_450 | NM24016, Pima 3-79 | 428 | CO088213 | *Gossypium raimondii* cDNA clone | 205/260 | 2e-57 |
| N3M2_200 | NM24016 | 184 | DT465216 | *G. hirsutum* cDNA clone | 125/157 | 2e-30 |
| N3M2_445 | NM24016, Pima 3-79 | 167 | No homology | | | |
| N3M2_230 | NM24016, Pima 3-79 | 227 | AF060639 | *G. barbadense* repetitive DNA sequence | 156/198 | 9e-39 |
| N2M1_350 | TM1, NM24016, Pima 3-79 | 331 | CAN61322 | *Vitis vinfera* hypothetical protein | 45/93 | 1e-19 |
| N2M1_450 | Acala 1517-99, NM24016 | 420 | BQ403000 | *G. arboreum* 7-10dpa fiber library cDNA clone | 122/153 | 4e-27 |
| N1M2_263 | Pima 3-79 | 183 | No homology | | | |
| N4M2_190 | Acala 1517-99, NM24016 | 104 | No homology | | | |

homology to three cotton cDNA sequences (Fig. 4 in supplementary material) and one repetitive DNA sequence, and one showed high homology to a hypothetical protein in *Vitis vinfera* (Table 5). ATG-AFLP fragment N1M2_450 is highly homologous to a putative calcium-binding protein in *G. hirsutum*. N2M1_350 is also significantly homologous to an *Oryza sativa* chromosome 11 PAC clone AC145775 (104/149, 2e-06). N2M1_450 fragment is highly homologous to a DNA fragment, UBC169-800 (Accession number AY570292) associated with a restorer gene $Rf_1$ (159/199, 1e-40). It is also significantly homologous to a microsatellite sequence, DQ908403 (79/104, 8e-06) and a putative retrotransposon copia-like sequence, EF457753 (123/174, 3e-05), in cotton. This fragment also shares significant similarities to a gene encoding alcohol dehydrogenase A.

In-silico analysis of AFLP fragments of ESTs in cotton

To confirm that *Nsi*I/*Mse*I combination indeed increases the number of gene fragments, EST sequences from Upland cotton deposited in GenBank were used for a comparative analysis. Of a total of 13 Mb sequences from 21,160 unigenes including 6,990 transcript assemblies (TA) and 14,170 singletons based on TIGR Plant Transcript Assemblies Release 1 (Aug. 15, 2005) for *G. hirsutum*, the traditional enzyme combination *Eco*RI/*Mse*I (GAATTC/TTAA) generated 6,204 AFLP fragments (29.3%), while the ATG-anchored restriction enzyme combination *Nsi*I/*Mse*I (ATGCAT/TTAA) generated 7,020 ATG-AFLP fragments (33.2%), a 13.2% increase. Based on Release 3 (Sept. 28, 2006) for *G. hirsutum* including 70,667 unigenes (with 24,797 TAs and 45,870 singletons) of a total of 63 Mb sequences, *Eco*RI/*Mse*I captured 22,237 fragments (31.5%), while *Nsi*I/*Mse*I generated 22,865 fragments (32.4%), a 2.8% increase. This indicates that the ATG-anchored AFLP does generate more cDNA-AFLP fragments than the traditional AFLP using *Eco*RI enzyme from cotton cDNAs.

The in-silico analysis using *Arabidopsis thaliana* gene sequences showed very similar results. Of all the gene sequences including 5-UTRs (untranslated regions), cDNA (with only exons), introns, and 3′-UTRs, *Eco*RI/*Mse*I produced 44,507 AFLP fragments; whereas *Nsi*I/*Mse*I generated 49,040 ATG-AFLP fragments (a 10.2% increase). The numbers of AFLP fragments from introns produced by the two enzyme combinations were 7,902 and 11,904, respectively. The same trend was seen for the numbers of AFLP fragments in 3′-UTRs (1,575 vs. 2,126). But the reverse was true for 5′-UTRs (1,038 vs. 287). This confirms a higher number of *Nsi*I/*Mse*I fragments than that of *Eco*RI/*Mse*I fragments in gene regions, suggesting a higher occurrence of ATGCAT sites than that of GAATTC sites. However, it appeared that *Nsi*I/*Mse*I generated even higher number of AFLP fragments (40,344) from intergenic regions than *Eco*RI/*Mse*I did (26,253).

An in-silico analysis was also conducted using a simple eukaryote, the yeast genome sequences provided by http://insilico.ehu.es/AFLP. *Eco*RI/*Mse*I generated 530, 1,060 and 736 AFLP fragments for chromosome I, III and VI, respectively, while *Nsi*I/*Mse*I generated 387, 868 and 552 ATG-AFLP fragments, respectively. The total number of fragments generated by *Nsi*I/*Mse*I is reduced by 22.3% (2,326 vs. 1,807).

## Discussion

With a better understanding of the general genome structures of higher organisms, primers derived from simple sequence repeats (SSRs), conserved regions of transposons, or retrotransposons were used in combination with random or AFLP primers to develop a number of modified marker

systems (Weising et al. 2005). Most of these markers represent random samples of the genome and have been used in various research areas. However, in the quest for genes responsible for evolutionary traits and plant phenotypes, functional markers from transcribed regions of the genome have recently gained more attention. Sequence-related amplified polymorphism (SRAP) (Li and Quiros 2001) and targeted region amplified polymorphism (TRAP) (Hu and Vick 2003) represented two recent successful attempts to target gene regions in a high-throughput fashion.

In our current study, to take advantages of the traditional AFLP techniques, we propose substitution of a restriction enzyme that recognizes a restriction site containing ATG, called ATG-anchored AFLP, i.e. ATG-AFLP analysis. Six *Nsi*I and eight *Mse*I primers generated a total of 1,780 ATG-AFLP fragments, of which 750 (42%, 15–16 per primer combination) were polymorphic among four genotypes from two cultivated cotton species (Upland cotton, *G. hirsutum* and Pima cotton, *G. barbadense*). The number of ATG-AFLP markers was sufficient to separate the four genotypes into two groups, consistent with their evolutionary and breeding history. Using a recombination inbred line (RIL) population, a total of 62 polymorphic ATG-AFLP markers amplified with 8 primers combinations were mapped to 19 linkage groups with known chromosome anchored simple sequence repeat (SSR) markers. This indicates that ATG-AFLP markers are widespread in the cotton genome.

It is known that ATG is the starting codon for most if not all of the protein-coding genes in higher plants and ATG is also located in UTRs and internally in gene sequences including exons and introns. ATG can also be found in non-coding regions including AT-rich sequences where ATG may expectedly occur more frequently. Otherwise, it may be distributed more or less randomly throughout the genome. Our in-silico analysis on *Arabidopsis thaliana* genome sequences demonstrated that ATGCAT/TTAA does generate more cDNA AFLP fragments than GAATTC/TTAA fragments; but higher percentage of ATG-AFLP fragments was also produced in intron and intergenic regions. However, the in-silico analysis of a less complex yeast genome indicated that ATG-AFLP fragments are reduced significantly. Of the nine ATG-AFLP fragments cloned and sequenced, three were found to be highly homologous to cotton cDNA sequences and one to a hypothetical protein, indicating almost half (44.4%) of the ATG-AFLP markers were amplified from gene regions for identification of candidate genes underlying quantitative trait variations. Even though a larger number of ATG-AFLP fragments should have been sequenced for verification, most of AFLP sequenced by others were reported to have low similarity to known gene sequences (Meudt and Clarke 2007; Yue et al. 2005). The increase in the number of gene-targeted fragments in ATG-AFLP analysis supports the notion that ATG-AFLP markers containing gene fragments are enriched.

The fact that *Nsi*I/*Mse*I yielded more cDNA-AFLP fragments than *Eco*RI/*Mse*I on cotton ESTs and *Arabidopsis* cDNAs suggests that ATG-anchored AFLP is also well suited for transcript profiling. In fact, *Nla*III and *Nco*I containing ATG in their restriction sites were used in cDNA-AFLP analysis in *Arabidopsis thaliana* and barley (Vuylsteke et al. 2007b). Except for *Nsi*I (ATGCAT), many other enzymes can also recognize sequences containing ATG, including *Nco*I (CCATGG), *Sph*I (GCATGC), *Pci*I (ACATGT), *Bsp*HI (TCATGA), *Pag*I (TCATGA), *Rca*I (TCATGA), *Cvi*AII (CATG), *Nla*III (CATG), *Fat*I (CATG), *Fok*I (GGATG(N)$_9$), *Btg*ZI (GCGATG(N)$_{10}$), and *BstF5*I (GGATGNN). These restriction enzymes can be selected in combination with the most often used enzymes such as *Eco*RI, *Mse*I, *Taq*I or others to perform ATG-AFLP analysis to ensure a genome-wide coverage in mapping, fingerprinting and transcript profiling for any organisms. Of course, with genome sequences available for several other plant species such as rice, poplar, maize, soybean, and sorghum, enzyme and primer combinations can be optimized by in-silico analysis. An in-depth in-silico analysis on the complete sequenced genomes should also shed light into the frequency of ATG and its distributions in gene regions and non-gene regions. Nevertheless, for the first time, the present study presents experimental evidence in plants using cotton as an example that ATG-AFLP targeting ATG in selection of restriction enzymes has advantages over the traditional AFLP analysis.

**Acknowledgments** This research was supported in part by grants from Cotton Incorporated, Cary, NC, USA, USDA-ARS, and the New Mexico Agricultural Experiment Station.

## References

Abdalla AM, Reddy OUK, El-Zik KM, Pepper AE (2001) Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. Theor Appl Genet 103:547–554

Altaf MK, Zhang JF, Stewart JM, Cantrell RG (1998) Integrated molecular map based on a trispecific F2 population of cotton. Proc Beltwide Cotton Conf, Natl Cotton Council Am, Memphis, pp 491–492

Bensch S, Akesson M (2005) Ten years of AFLP in ecology and evolution: why so few animals? Mol Ecol 14:2899–2914

Bikandi J, San Millán R, Rementeria A, Garaizar J (2004) In silico analysis of complete bacterial genomes: PCR, AFLP-PCR, and endonuclease restriction. Bioinformatics 20:798–799

Brubaker CL, Brown AH (2003) The use of multiple alien chromosome addition aneuploids facilitates genetic linkage mapping of the *Gossypium* G genome. Genome 46:774–791

Cantrell RG, Davis DD (2000) Registration of NM24016, an interspecific-derived cotton genetic stock. Crop Sci 40:1208

Cantrell RG, Roberts CL, Waddell C (2000) Registration of 'Acala 1517–99'cotton. Crop Sci 40:1200–1201

Childs KL, Hamilton JP, Zhu W, Ly E, Cheung F, Wu H, Rabinowicz PD, Town CD, Buell CR, Chan AP (2007) The TIGR plant transcript assemblies database. Nucleic Acids Res 35:D846–D851

Endrizzi JE, Turcotte EL, Kohel RJ (1984) Qualitative genetics, cytology, and cytogenetics. In: Kohel RJ, Lewis CF (eds) Cotton. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America. Madison, Wisconsin pp 82–129, 605 p

Hampl V, Pavlícek A, Flegr J (2001) Construction and bootstrap analysis of DNA fingerprinting-based phylogenetic trees with a freeware program FreeTree: application to trichomonad parasites. Int J Syst Evol Microbiol 51:731–735

Hu J, Vick BA (2003) TRAP (target region amplification polymorphism), a novel marker technique for plant genotyping. Plant Mol Biol Rep 21:289–294

Iqbal MJ, Reddy OUK, El-Zik KM, Pepper AE (2001) A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. Theor Appl Genet 103:547–554

Kohel RJ, Richmond TR, Lewis CF (1970) Texas Marker-1. Description of a genetic standard for *Gossypium hirsutum* L. Crop Sci 10:670–671

Lacape JM, Nguyen TB, Thibivilliers S, Bojinov B, Courtois B, Cantrell RG, Burr B, Hau B (2003) A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* x *Gossypium barbadense* backcross population. Genome 46:612–626

Lacape JM, Nguyen TB, Courtois B, Belot JL, Giband M, Gourlot JP, Gawryziak G, Roques S, Hau B (2005) QTL analysis of cotton fiber quality using multiple *Gossypium hirsutum* x *Gossypium barbadense* backcross generations. Crop Sci 45:123–140

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Li G, Quiros CF (2001) Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: Its application to mapping and gene tagging in *Brassica*. Theor Appl Genet 103:455–461

Liu B, Brubaker CL, Mergeai G, Cronn RC, Wendel JF (2001) Polyploid formation in cotton is not accompanied by rapid genomic changes. Genome 44:321–330

Mei M, Syed NH, Gao W, Thaxton PM, Smith CW, Stelly DM, Chen ZJ (2004) Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). Theor Appl Genet 108:280–291

Meudt HM, Clarke AC (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances. Trends Plant Sci 12:106–117

Niu C, Hinchliffe DJ, Cantrell RG, Wang C, Roberts PA, Zhang JF (2007) Identification of molecular markers associated with root-knot nematode resistance in Upland cotton. Crop Sci 47:951–960

Park KC, Kim NH, Cho YS, Kang KH, Lee JK, Kim NS (2003) Genetic variations of AA genome *Oryza* species measured by MITE-AFLP. Theor Appl Genet 107:203–209

Pillay M, Myers GO (1999) Genetic diversity in cotton assessed by variation in ribosomal RNA genes and AFLP markers. Crop Sci 39:1881–1886

Rombauts S, Van de Peer Y, Rouz P (2002/3) AFLPinSilico, simulating AFLP fingerprints. Bioinformatics 19:776–777

Roy JK, Balyan HS, Prasad M, Gupta PK (2002) Use of SAMPL for a study of DNA polymorphism, genetic diversity and possible gene tagging in bread wheat. Theor Appl Genet 104:465–472

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Vuylsteke M, Peleman JD, van Eijk MJ (2007a) AFLP technology for DNA fingerprinting. Nat Protoc 2:1387–1398

Vuylsteke M, Peleman JD, van Eijk MJ (2007b) AFLP-based transcript profiling (cDNA-AFLP) for genome-wide expression analysis. Nat Protoc 2:1399–1413

Wang F, Stewart JM, Zhang JF (2007) Molecular markers linked to the $Rf_2$ fertility restorer gene in cotton. Genome 50:818–824

Weising K, Nybom H, Wolff K, Kahl G (2005) DNA fingerprinting in plants: principles, methods, and applications, 2nd edn. CRC Press, Boca Raton, FL

Westengen OT, Huaman Z, Heun M (2005) Genetic diversity and geographic pattern in early South American cotton domestication. Theor Appl Genet 110:392–402

Van der Wurff AWG, Chan YL, van Straalen NM, Schouten J (2000) TE-AFLP: combining rapidity and robustness in DNA fingerprinting. Nucleic Acids Res 28:e105

Yue Z, Wang W, Kong J, Dai J (2005) Isolation, cloning, and sequencing of AFLP markers related to disease resistance traits in *Fennerospenaeus chinensis*. Chin J Oceanol Limnol 23:442–447

Zhang JF, Stewart JM (2000) Economical and rapid method for extracting cotton genomic DNA. J Cotton Sci 4:193–201

Zhang JF, Lu YZ, Yu SX (2005a) Cleaved AFLP (cAFLP), a modified amplified fragment length polymorphism analysis for cotton. Theor Appl Genet 111:1385–1395

Zhang JF, Lu Y, Percy RG, Ulloa M, Becelaere G, Chee P, Cantrell RG (2005b) A molecular linkage map and quantitative trait locus analysis based on a recombinant inbred line population of cotton. Proc Beltwide Cotton Conf, Natl Cotton Council Am, Memphis, p 899

Zhang JF, Yuan YL, Niu C, Hinchliffe DJ, Lu Y, Yu S, Percy RG, Cantrell RG (2007) AFLP-RGA markers in comparison with RGA and AFLP in cultivated tetraploid cotton. Crop Sci 47:180–187